SA1

Dizon, Kiana Marie

2024-10-12

1. Hypothesis Formulation and Rationale (10 points)

We want to determine if there is a significant difference in the mean time spent on cell phones between male and female students per week. The hypothesis can be formulated as:

- Null Hypothesis (H_0): The mean time spent on cell phones by males and females is the same, $\mu_1 = \mu_2$.
- Alternative Hypothesis (H_1): The mean time spent on cell phones by males and females is different, $\mu_1 \neq \mu_2$.

We will use a two-sample t-test to compare the means because the two samples (males and females) are independent.

2. Hypothesis Test Conclusion and p-value (30 points)

We perform a two-sample t-test on the data provided to determine if the means are significantly different.

```
males <- c(12, 4, 11, 13, 11, 11, 7, 9, 10, 7, 10, 10, 9, 6, 9, 15, 11, 12, 7, 8, 10, 11, 12, 7, 8, 9, 8, 11, 9,
11, 10, 9, 9, 7, 7, 8, 11, 12, 8, 8, 12, 8, 11, 10, 10, 10, 8, 10, 13, 13)
females <- c(9, 7, 10, 9, 7, 10, 7, 10, 7, 9, 9, 6, 11, 8, 9, 9, 12, 14, 12, 12, 8, 12, 9, 10, 11, 7, 9, 8, 11,
9, 12, 13, 8, 13, 8, 10, 11, 9, 11, 7, 9, 11, 8, 9, 11, 9, 12, 11, 8, 11)

t_test_result <- t.test(males, females, alternative = "two.sided", var.equal = FALSE)

t_test_result
```

```
##
##  Welch Two Sample t-test
##
## data:  males and females
## t = 0, df = 96.545, p-value = 1
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.7929637  0.7929637
## sample estimates:
## mean of x mean of y
##      9.64      9.64
```

Hypothesis Test Conclusion and p-value The test provides a p-value which helps us to draw a conclusion:

- If the p-value is less than the significance level (usually 0.05), we reject the null hypothesis.
- If the p-value is greater than 0.05, we fail to reject the null hypothesis.

Recommendation: Based on the p-value, we will provide a recommendation for the researcher whether there is a significant difference between the mean time spent by males and females on cell phones.

3. Descriptive Statistical Summaries (10 points)

We calculate the mean, standard deviation, and sample size for each gender.

```
mean_males <- mean(males)
mean_females <- mean(females)
sd_males <- sd(males)
sd_females <- sd(females)
n_males <- length(males)
n_females <- length(females)

summary_statistics <- data.frame(
  Gender = c("Males", "Females"),
  Mean = c(mean_males, mean_females),
  SD = c(sd_males, sd_females),
  N = c(n_males, n_females)
)

summary_statistics
```

```
##    Gender Mean       SD  N
## 1   Males 9.64 2.116601 50
## 2 Females 9.64 1.870938 50
```

4. 95% Confidence Intervals (10 points)

We calculate the 95% confidence intervals for the mean of each gender and the difference in means between the two groups.

```
ci_males <- t.test(males)$conf.int

ci_females <- t.test(females)$conf.int

ci_diff <- t.test(males, females)$conf.int

list(ci_males = ci_males, ci_females = ci_females, ci_diff = ci_diff)
```

```
## $ci_males
## [1]  9.038469 10.241531
## attr(,"conf.level")
## [1] 0.95
##
## $ci_females
## [1]  9.108285 10.171715
## attr(,"conf.level")
## [1] 0.95
##
## $ci_diff
## [1] -0.7929637  0.7929637
## attr(,"conf.level")
## [1] 0.95
```

5. Discussion on Sample Size and Further Testing (20 points)

The current sample size of 50 males and 50 females provides a decent basis for this analysis. However, larger sample sizes can increase the reliability of the results and provide more generalizability. The variability in cell phone usage may differ across different age groups or regions, so additional testing with more diverse samples would be helpful to confirm the findings.

6. Assumptions for Two-Sample T-Test (20 points)

We check two assumptions for the two-sample t-test:

Normality: The Shapiro-Wilk test is used to check if the data follows a normal distribution.

Homogeneity of Variance: We use the F-test to check if the variances are equal between the two groups.

Normality Test

```
shapiro_males <- shapiro.test(males)
shapiro_females <- shapiro.test(females)

shapiro_males
```

```
##
##  Shapiro-Wilk normality test
##
## data:  males
## W = 0.97572, p-value = 0.3887
```

```
shapiro_females
```

```
##
##  Shapiro-Wilk normality test
##
## data:  females
## W = 0.95748, p-value = 0.06972
```

Homogeneity of Variance Test

```
var_test <- var.test(males, females)

var_test
```

```
##
##  F test to compare two variances
##
## data:  males and females
## F = 1.2799, num df = 49, denom df = 49, p-value = 0.3908
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.7262848 2.2553383
## sample estimates:
## ratio of variances
##           1.279851
```

Conclusion on Assumptions

Based on the results of the normality test and the F-test for variances, we conclude whether the assumptions of the t-test are met. If the assumptions are not met, other non-parametric tests such as the Mann-Whitney U test may be considered.